UECM1703 Introduction of Scientific Computing

TOPIC 6 Practical

UNIVERSITI TUNKU ABDUL RAHMAN

Faculty:	FES	Unit Code:	UECM1703
Course:	AM &FM	Unit Title:	Introduction To Scientific Computing
Year:	1&2	Lecturer:	Dr Yong Chin Khian
Session:	Oct 2022		

Q1. A local Bank wants to estimate the number of foreclosures per 1000 houses sold. Bank officials think that the number of foreclosures is related to the size of the down payment made by house buyers. The following table contains foreclosure and down payment data.

Down Payment Size (% of purchase price)	Number of Foreclosures per 1000 houses
10	42
20	11
14	29
12	25
18	12
16	24

You fit the above data to $y = \beta_0 + \beta_1 x + \epsilon$, where y is the number of foreclosures per 1000 houses sold, and x is the size of the down payment made by house buyers. Use the stat.models module to compute the adjusted R^2.

Q2. Consider the data shown below:

y	x	y	x
1791	16	971	13
450	10	594	11
594	11	1791	16
331	9	103	6
1207	14	594	11
103	6	594	11
2143	17	1480	15
450	10	331	9

Fit the polynomial regression models up to order 4. A model that best fit the data based on the AIC criteria was fitted, determine the fitted value of y when $x = 10$.

Q3. Consider the data shown below:

y	x	y	x
1860	16	1006	13
464	10	615	11
615	11	1860	16
341	9	105	6
1252	14	615	11
105	6	615	11
2226	17	1536	15
464	10	341	9

Fit the polynomial regression models up to order 4. A model that best fit the data based on the BIC criteria was fitted, determine the corresponding BIC values.

Q4. Consider the data shown below:

y	24	25	19	29	20	23	26	21	21	23	26	31
x	12	13	8	17	9	11	14	10	10	11	14	19

You fit the above data to $Y = \beta_0 + \beta_1 X + \epsilon$. Obtain a prediction for a new observation y_h when $x_h = 18.0$.

Q5. You fit the following data to $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$ with independent normal error terms.

y	x_1	x_2	x_3	y	x_1	x_2	x_3
105.6	7.9	46.2	15.4	86.8	4.6	23.9	8.7
87.4	4.7	24.7	8.9	82.9	3.9	19.3	7.3
99.2	6.8	38.7	13.1	72.4	2.0	7.0	3.6
106.5	8.1	47.4	15.7	85.5	4.4	22.4	8.2
94.7	6.0	33.2	11.5	89.8	5.1	27.5	9.8
104.3	7.7	44.6	14.9	74.5	2.4	9.4	4.3
89.1	5.0	26.8	9.5	88.6	4.9	26.2	9.3
101.5	7.2	41.3	13.9	82.7	3.9	19.0	7.2

Use the stat.models package in Python to obtain a prediction for a new observation Y_h when $x_{h_1} = 5.9$, $x_{h_2} = 30.8$ and $x_{h_3} = 10.2$.

Q6. You fit $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$ with independent normal error terms to data in data file dfQ09c.csv. Use the stat.models package in Python to answer the following questions.

(a) State the estimated regression function.

(b) Obtain a prediction for a new observation Y_h when $x_{h_1} = 5.9$, $x_{h_2} = 30.8$, $x_{h_3} = 10.2$ and $x_{h_4} = 19.0$.

(c) Write down the python codes and the correspoding outputs that you obtained for answers in part (a) and part (b).

UECM1703 Introduction of Scientific Computing

Q7. The quality of orange juice produced by a manufacturer is constantly monitored. Is there a relationship between the sweetness index and chemical measure such as the amount of water soluble pectin (parts per million) in the orange juice? Data collected on these two variables for 24 production runs at juice manufacturing plant are save in orjuice.csv file. Suppose a manufacturer wants to use simple linear regression to predict the sweetness(Y) from the amount of pectin (X).

(a)Fit the data to the model $Y = \beta_0 + \beta_1 X + \epsilon$.

(b)Plot the fitted regression function and the data.

Q8. Consider the data shown below:

y	x	y	x
1008246	32	377998	25
155960	20	227586	22
155960	20	888652	31
81938	17	20711	12
592936	28	189241	21
10111	10	189241	21
1139524	33	681678	29
127274	19	81938	17

Fit the polynomial regression models up to order 4, then answer the following questions:

(a)Fill in the table below:

Polynomial model	Adjusted R^2	AIC	BIC
First order			
Second order			
Third order			
Fourth order			

(b)Determine a model that best fit the data based on the adjusted R^2.

(c)Write down the equation of the best fitted curve that you have selected from part (b).

(d)Created a scatter plot of the data and add the fitted polynomial line to the scatterplot.

Q9. You are given the following data:

x	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
y	3	14	23	25	23	15	9	5	9	13	17	24	32	36	46

Fit the data to the polynomial regression model of degree 4. Created a scatter plot of the data and add the fitted polynomial line to scatterplot.

Q10. Consider the data below:

y	x_1	x_2	x_3	x_4
268.6	7.9	31.6	15.4	30.2
181.7	4.6	29.4	8.7	16.8
184.9	4.7	29.5	8.9	17.3
164.1	3.9	28.9	7.3	14.1
239.4	6.8	30.9	13.1	25.7
116.0	2.0	27.7	3.6	6.7
273.7	8.1	31.7	15.7	31.0
175.8	4.4	29.2	8.2	15.9
218.0	6.0	30.3	11.5	22.4
195.9	5.1	29.8	9.8	19.0

Assuming that regression model $y = \beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+\epsilon$ is appropriate. Find the test statistic for testing $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$.

Q11. Consider the data below:

y	x_1	x_2	x_3	x_4
223.5	5.6	30.1	10.7	20.8
242.9	6.2	30.5	12.0	23.5
177.8	4.0	29.0	7.5	14.5
251.6	6.5	30.7	12.6	24.7
203.1	4.9	29.6	9.2	18.0
152.3	3.1	28.4	5.7	11.0
193.0	4.5	29.4	8.6	16.6
247.3	6.4	30.6	12.3	24.1
235.7	6.0	30.3	11.5	22.5
171.4	3.8	28.9	7.1	13.6

Assuming that regression model $y = \beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+\epsilon$ is appropriate. Find the test statistic for testing $H_0 : \beta_2 = 0$.

Q12. Consider the data below:

y	x_1	x_2	x_3	x_4
206.6	5.6	30.1	10.7	20.8
223.7	6.2	30.5	12.0	23.5
166.0	4.0	29.0	7.5	14.5
231.4	6.5	30.7	12.6	24.7
188.5	4.9	29.6	9.2	18.0
143.5	3.1	28.4	5.7	11.0
179.6	4.5	29.4	8.6	16.6
227.7	6.4	30.6	12.3	24.1
217.4	6.0	30.3	11.5	22.5
160.4	3.8	28.9	7.1	13.6

Assuming that regression model $y = \beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_4x_4+\epsilon$ is appropriate. Find the p-value for testing $H_0 : \beta_2 = 0$.

Q13. Consider the data shown below:

y	x	y	x
1065	11.2	660	10.1
449	9.3	521	9.6
449	9.3	1022	11.1
366	8.9	222	8.0
826	10.6	496	9.5
197	7.8	496	9.5
1109	11.3	901	10.8
406	9.1	347	8.8

You fit the above data to $y = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3 + \beta_4x^4 + \epsilon$.

(a) Write down the equation of the fitted curve.

(b) s the model $y = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3x^3 + \beta_4x^4 + \epsilon$ significant for estimating y? Justify your answer using p-value.

(c) Determine the predicted value of the mean of y when $x = 7.7$.

(d) Created a scatter plot of the data with the fitted polynomial line added to the scatterplot.